CCseqBasic - system requirements , and LINUX environment

SERVER ADDRESS AND PUBLIC ACCESS DATA AREA

The pipeline needs a server address, and a publicly visible data area served via this server. It is not enough to be visible “inside your institute” : UCSC genome browser has to see the generated data hubs, and the genome browser is not located inside your institute !

Your command line tools should be allowed to write into this publicly visible disk area (i.e. wherever you run the pipeline, you should have the public area visible and write-able).

Ask your system admin / IT support about this.

TOOLKITS THE PIPELINE NEEDS TO RUN

TABLE 1 - list of tools CCseqBasic needs to run.

Toolkit	Tested with	Supports	Does not support
samtools	1.1	1.x	0.x
bowtie	1.1.2	0.x , 1.x	
bowtie2	2.1.0	2.x	
bedtools	2.17.0	2.1x	2.2x
flash	1.2.8	1.x	?
trim_galore	0.3.1	0.x	?
cutadapt	1.2.1	1.x	?
fastqc	0.11.4	0.11.x	0.10.x
multiqc	0.7	?	?
blat	35	?	?
perl	5.18.1	5.x	?
python	2.7.5	2.x	3.x
matplotlib (in python)	1.4.1	1.4.x -->	1.3.x or older

Multiqc is only needed for CCseqBasic5_rainbow and CCseqBasic5_parallel and newer code versions.

In addition, the pipeline contains wigToBigWig and bedToBigBed from ucsctools
(downloaded from http://hgdownload.soe.ucsc.edu/admin/exe/)

BOWTIE GENOME INDICES

The pipeline requires already-made bowtie genome indices (for the genome builds you are interested in), built with UCSC genome build fasta.

If you have ever mapped your data using bowtie (using UCSC, not Ensembl coordinates for your chromosomes) you will already have these.

Downloadable files of the Illumina iGenomes UCSC build bowtie1 indices are available :

Bowtie main web site (right hand side panel “pre-built indices)
http://bowtie-bio.sourceforge.net/index.shtml

iGenomes web site (each download contain ready-made both bowtie1 and bowtie2 indices)
https://support.illumina.com/sequencing/sequencing_software/igenome.html
If you want to use UCSC-supported FASTA, for which there is no iGenomes index,
you can build your own index :

UCSC - supported FASTAs to feed into bowtie index builder :
http://hgdownload.cse.ucsc.edu/downloads.html

Bowtie1 - build your indices :
http://bowtie-bio.sourceforge.net/manual.shtml#the-bowtie-build-indexer

For genomes not already supported by the pipe, you need to also provide UCSC chromosome size file
You can do it like this : fetchChromSizes mm9 > mm9.sizes.txt (fetches the mm9 build chromSizes file).
The fetchChromSizes is part of UCSCtools toolkit, individual tools downloadable at :
http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/

SYSTEM SPECIFICATIONS

(within the system where CCseqBasic was developed and tested)

GNU basic tool versions

(GNU coreutils) 8.4 - Copyright (C) 2010 Free Software Foundation, Inc.
GNU which v2.19 - Copyright (C) 1999 - 2008 Carlo Wood.
GNU Awk 3.1.7 - Copyright (C) 1989, 1991-2009 Free Software Foundation.
GNU sed version 4.2.1 - Copyright (C) 2009 Free Software Foundation, Inc.
grep (GNU grep) 2.20 - Copyright (C) 2014 Free Software Foundation, Inc.

bc : the unix basic calculator (forgot to check which version this is)

Other BASH utils (not obligatory) :

hostname - hostname 1.100 (2001-04-14), net-tools 1.60
module - VERSION=3.2.10 , DATE=2012-12-21

Perl version

This is perl 5, version 18, subversion 1 (v5.18.1) built for x86_64-linux-thread-multi
Copyright 1987-2013, Larry Wall

Python version

Python 2.7.5

Machines where the pipe was tested : (uname -a)

```
Linux 2.6.32-642.11.1.el6.x86_64 #1 SMP x86_64 x86_64 x86_64 GNU/Linux
Linux 2.6.32-642.6.2.el6.x86_64  #1 SMP x86_64 x86_64 x86_64 GNU/Linux
```

Sun grid engine version

GE 6.2u5